



Sequence variation in the $\beta 7$ – $\beta 8$ loop of bacterial class A sortase enzymes alters substrate selectivity

Received for publication, March 15, 2021, and in revised form, July 9, 2021. Published, Papers in Press, July 22, 2021, <https://doi.org/10.1016/j.jbc.2021.100981>

Isabel M. Piper^{1,‡}, Sarah A. Struyvenberg^{1,‡}, Jordan D. Valgardson¹, D. Alex Johnson¹, Melody Gao¹, Katherine Johnston¹, Justin E. Svendsen¹, Hanna M. Kodama¹, Kelli L. Hvorecny², John M. Antos^{1,*}, and Jeanine F. Amacher^{1,*}

From the ¹Department of Chemistry, Western Washington University, Bellingham, Washington, USA; and ²Department of Biochemistry, University of Washington, Seattle, Washington, USA

Edited by Joseph Jez

Gram-positive bacteria contain sortase enzymes on their cell surfaces that catalyze transpeptidation reactions critical for proper cellular function. *In vitro*, sortases are used in sortase-mediated ligation (SML) reactions for a variety of protein engineering applications. Historically, sortase A from *Staphylococcus aureus* (saSrtA) has been the enzyme of choice to catalyze SML reactions. However, the stringent specificity of saSrtA for the LPXTG sequence motif limits its uses. Here, we describe the impact on substrate selectivity of a structurally conserved loop with a high degree of sequence variability in all classes of sortases. We investigate the contribution of this $\beta 7$ – $\beta 8$ loop by designing and testing chimeric sortase enzymes. Our chimeras utilize natural sequence variation of class A sortases from eight species engineered into the SrtA sequence from *Streptococcus pneumoniae*. While some of these chimeric enzymes mimic the activity and selectivity of the WT protein from which the loop sequence was derived (e.g., that of saSrtA), others results in chimeric *Streptococcus pneumoniae* SrtA enzymes that are able to accommodate a range of residues in the final position of the substrate motif (LPXTX). Using mutagenesis, structural comparisons, and sequence analyses, we identify three interactions facilitated by $\beta 7$ – $\beta 8$ loop residues that appear to be broadly conserved or converged upon in class A sortase enzymes. These studies provide the foundation for a deeper understanding of sortase target selectivity and can expand the sortase toolbox for future SML applications.

Sortases are cysteine transpeptidase enzymes that gram-positive bacteria use to covalently attach proteins to their cell wall for various functions, including the assembly of pili or display of virulence factors (1–3). There are six recognized classes of sortase enzymes (classes A–F), with roles *in vivo* ranging from general purpose or “housekeeping” functions (classes A and E), to more specific roles such as the

construction of the bacterial pilus (class C) (1, 4). These enzymes recognize a cell wall sorting signal (CWSS) on the outer membrane of gram-positive bacteria (1, 5). For class A sortases, the CWSS is the sequence LPXTG (1, 5). Using previously published numbering (L = P4, P = P3, X = P2, T = P1, and G = P1'), P4, P3, and/or P1' of this motif vary among different classes (5). After target recognition, a His–Cys–Arg catalytic triad facilitates a transpeptidation reaction whereby the CWSS is first cleaved between the P1 and P1' residues *via* nucleophilic attack by the catalytic Cys, resulting in a thioester linkage with the P1 position of the CWSS. Resolution of this acyl-enzyme intermediate is then achieved by nucleophilic attack by an amino group displayed on the cell wall building block lipid II, or in the case of pilus formation, displayed on a separate protein subunit (1, 3, 5, 6). The final result is the formation of a new amide linkage, with the portion of the substrate N-terminal to the CWSS now covalently attached at its C terminus to the amine nucleophile ligation partner.

The ability to cleave a substrate sequence and subsequently ligate a second component (for example a protein or synthetic peptide derivative) makes sortases an attractive tool for protein engineering efforts, commonly called sortase-mediated ligation (SML) or sortagging (3). Sortase A from *Staphylococcus aureus* (saSrtA) was the first of these enzymes discovered and continues to see widespread use for *in vitro* SML experiments (1, 7). Notable improvements in SML technology have occurred in recent years, including strategies for limiting the reversibility of the ligation reaction and the development of saSrtA variants with dramatically improved catalytic efficiency (3, 8, 9). However, as a consequence of the narrow substrate selectivity of saSrtA (10), the majority of SML examples rely on the combination of one ligation partner displaying an LPXTG motif near its C-terminus with another possessing one or more N-terminal glycines. This restricted substrate scope can be advantageous, for example, in the use of SML for labeling specific polypeptides in complex mixtures, but it also represents a limitation for certain applications (9, 11, 12). Highlighting this point, an increasing number of studies have demonstrated that the use of naturally occurring sortases or engineered sortases with altered substrate selectivity offers distinct advantages such as reducing

[‡] These authors contributed equally to this work.

* For correspondence: Jeanine F. Amacher, amachej@wwu.edu; John M. Antos, antosj@wwu.edu.

Present address for Jordan D. Valgardson: Department of Chemical Systems Biology, Stanford University, Palo Alto, CA, USA.

Selectivity determinants in class A sortases

the necessity for point mutations in protein semisynthesis applications (12), enabling the labeling of endogenous proteins that do not naturally contain the LPXTG motif (11, 13), and allowing labeling of multiple sites within the same protein target (11, 14). Thus, the engineering and discovery of sortases with altered substrate profiles, along with a better understanding of the biochemical basis for sortase substrate selectivity, represent important areas for the continued development of SML technology.

Previous mutagenesis and structural studies of various sortases provide a wealth of knowledge about substrate recognition, including initial ligand recognition and subsequent cleavage (thioesterification), as well as nucleophile recognition and mechanistic details of peptide ligation (transpeptidation) (1, 2, 15). Specifically, the catalytic residues of all native sortases identified to date are (using saSrtA numbering unless specified otherwise) as follows: His120 (general acid/base), Cys184 (nucleophile, acyl-enzyme intermediate), and Arg197 (transition state stabilization) (Fig. 1A) (1, 15). In addition, directed evolution studies have identified mutations (P94R/D160N/D165A/K190E/K196T) that are together able to boost the catalytic efficiency of saSrtA by 120-fold (8). Of these five mutations, several are in two of the three structurally conserved loops in class A sortases located near

the peptide-binding cleft: those between the β 4, β 5 strands (β 4– β 5 loop), the β 6, β 7 strands (β 6– β 7 loop, where D165A occurs), and the β 7, β 8 strands (β 7– β 8 loop, where K190E and K196T are located). Notably, while the increase in enzyme activity afforded by these mutations included a 3.6-fold increase in k_{cat} , the effect was dominated by a 33-fold decrease in K_M , suggesting these loop residues may be important in CWSS recognition (8).

Additional evidence for the role of loop residues was obtained from more targeted directed evolution and mutagenesis studies. For example, it was demonstrated that the β 6– β 7 loop of saSrtA directly confers specificity at P4 of the recognition motif (LPXTG), and residues other than leucine (L) can be accommodated using sortases with mutations in the β 6– β 7 loop (12, 16, 17). Indeed, substitution of the β 6– β 7 loop residues from saSrtB into the saSrtA enzyme alters substrate recognition to that of a sortase B protein (NPQTN) (18). Turning to the β 7– β 8 loop, the NMR structure of saSrtA covalently bound to a modified LPAT* peptide mimetic revealed a noncovalent interaction between W194 in saSrtA and the Thr residue in P1 (LPXTG) (19, 20). Mutation of W194 in saSrtA decreased the reaction rate, although it was not essential to catalysis (20). Taken together, these past studies reveal that sequence variation within sortase loops

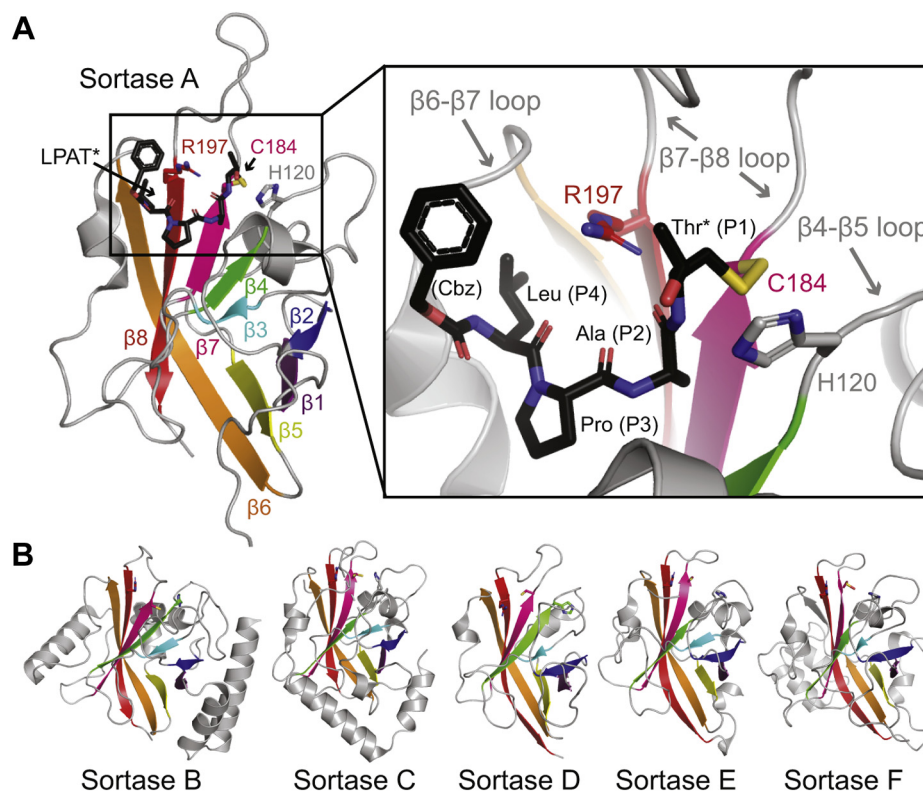


Figure 1. The sortase-fold is conserved in all classes of bacterial sortases. A, the peptide-bound structure of *S. aureus* SrtA (saSrtA) is shown in cartoon representation, with β -strands colored and labeled (PDB ID: 2KID) (19). The side chains of the catalytic residues (H120, C184, and R197) are shown as sticks, colored by heteroatom (O = red, N = blue, S = yellow), and labeled. The disulfide-linked peptide analog, Cbz-LPAT*, where Cbz is a carbobenzyloxy protecting group and T* is (2*R*,3*S*)-3-amino-4-mercapto-2-butanol, is shown as black sticks and colored by heteroatom (19). A zoomed-in version of the active site is shown in the black box, with features indicated as in panel A. The variable loops are labeled and indicated by gray arrows. B, the overall sortase fold is well conserved in proteins of different classes. Here, structures for class B (PDB ID: 1NG5), class C (3O0P), class D (2LN7), class E (5CUW), and class F (5UUS) sortases are shown in cartoons, with conserved β -strands colored as in panel A, highlighting the 8-stranded sortase fold. The conserved catalytic triad is shown in sticks (and colored by heteroatom) for all.

directly affects both activity and selectivity for target ligands. Furthermore, conservation of the closed eight-stranded β -barrel core in all sortase A-F structures that have been reported to date suggests that these principles may apply to non-class A sortases as well (Fig. 1B) (2).

In this work, we specifically look at natural sequence variation in the β 7– β 8 loop of class A sortases, using *Streptococcus pneumoniae* SrtA (spSrtA) as a model system. The β 7– β 8 loop was initially identified after sequence and evolutionary conservation analyses as a region of notable variability in class A (and other) sortases. We find that the β 7– β 8 loop sequence dramatically affects both overall enzyme activity and selectivity at P1' of the CWSS. Our data are consistent with a recent publication that investigated the grafting of β 7– β 8 loop sequences from saSrtA and *Bacillus anthracis* SrtA (baSrtA) into *Streptococcus pyogenes* SrtA (18). This work also suggested that W194 (saSrtA numbering) may play a role in the substrate recognition of the reported chimeras (21). Here, we have profiled the substrate preferences of over a dozen loop chimeras and single- or double-mutants targeting the β 7– β 8 loop. While we also observe a role for W194 in substrate recognition, our data suggest that it is unique to saSrtA and not broadly applicable to describe β 7– β 8 loop-mediated class A sortase function. Indeed, the combination of functional enzyme assays and analysis of reported sortase structures in the present work suggests three different β 7– β 8 loop-mediated interactions that affect selectivity and activity.

Results

Sequence analyses of bacterial sortases

To investigate general sequence variation in class A sortases, we first created a multiple sequence alignment of the eight SrtA sequences used in our previous work, proteins from: *B. anthracis*, *Enterococcus faecalis*, *Lactococcus lactis*, *Listeria monocytogenes*, *S. aureus*, *Streptococcus oralis*, *S. pneumoniae*, and *Streptococcus suis* (Fig. 2A) (22). Although this is a small subset of SrtA sequences, we reasoned that this representative group would reveal general sequence trends or sequence variations as these enzymes had exhibited clear differences in substrate preferences in our previous study (19). Indeed, variations were present in loop regions, specifically those of the β 2– β 3, β 4– β 5, β 6– β 7, and β 7– β 8 loops (Fig. 2A).

In addition to our sequence analysis, we also wanted to look more broadly at global evolutionary conservation in class A sortases. We extracted 400 SrtA sequences from the NCBI and used MAFFT to create a multiple sequence alignment (23). To visualize evolutionary conservation, we used the ConSurf server, with saSrtA (PDB ID: 2KID) as our structural template (Fig. 2B) (24, 25). This analysis confirmed that class A sortase sequences are quite variable, with very few residues showing a high degree of conservation (dark maroon color). However, validating our analysis, the peptide-binding cleft (occupied by the peptidomimetic LPAT* in 2KID) is very well conserved, as is the presumed endogenous lipid II-binding site on the backside of the enzyme (Fig. 2B) (2, 26).

As suggested by our initial multiple sequence alignment (Fig. 2A), ConSurf confirmed that the three structural loops that border the peptide binding cleft, the β 4– β 5, β 6– β 7, and β 7– β 8 loops, are all quite variable (Fig. 2C). Any conservation in these loops that we do see may also be an artifact of the multiple sequence alignment because of variable lengths; for example, E195 saSrtA in the β 7– β 8 loop appears to be highly conserved, but this is because of the fact that out of the 400 sequences used in the alignment, only 10 have any residue at this position (9 of which are Glu) and all are of *Staphylococcus* sequences.

Given that the β 6– β 7 loop has been shown to be intimately involved in sortase substrate recognition, we were intrigued that our analysis revealed similar levels of variability in the β 4– β 5 and β 7– β 8 loops (18). In the case of β 7– β 8, we were also motivated by its proximity to the enzyme active site and CWSS P1' position (15). Therefore, we sought to further explore how the β 7– β 8 loop affects the activity and substrate specificity of a sortase with narrow substrate tolerance (saSrtA) versus one that is more promiscuous (spSrtA).

Loop-swapped β 7– β 8 variants reveal differences in position P1' selectivity for *S. aureus* and *S. pneumoniae* SrtA enzymes

In our previous work, we found that the most striking differences in substrate tolerances among the class A sortases studied were observed at the P1' (LPXTG) of the substrate motif (19). For example, while saSrtA is specific for a Gly residue at P1', SrtA from *S. pneumoniae* (spSrtA) recognizes over 10 of the 20 amino acids at this position in a 24-h endpoint assay (22). To determine whether the β 7– β 8 loop played a role in these differing substrate preferences, we began by engineering two loop-swapped variants: saSrtA_{pneumoniae} (which contains the β 7– β 8 loop residues from spSrtA (CEDLAATER, where the catalytic cysteine and arginine are underlined)) and spSrtA_{aureus} (with β 7– β 8 residues CDDYNEKTGVWEKR from saSrtA). Notably, the length of the saSrtA β 7– β 8 loop contains an additional five residues, as compared with the spSrtA β 7– β 8 loop. The saSrtA loop also uniquely contains W194, which is predicted to directly contact the P1 threonine of the LPXTG motif (19). In addition, while both loops are predicted to have an overall net negative charge at the physiological pH, the saSrtA loop contains two positively charged lysine residues that are not present in spSrtA. Both chimeric sortases were expressed and purified from *Escherichia coli* and were isolated as soluble, monomeric enzymes as described previously and in Experimental procedures (Fig. S1) (22). Based on the migration of these variants using size-exclusion chromatography (SEC), the variants are not aggregated and retain a similar radius of gyration and oligomeric status (Fig. S1).

To monitor enzymatic activity and selectivity of the saSrtA_{pneumoniae}, spSrtA_{aureus}, and their WT counterpart proteins, we utilized well-established FRET quencher probes consisting of different substrate motifs flanked by a 2-aminobenzoyl fluorophore (Abz) and a 2,4-dinitrophenyl quencher (Dnp) (20, 27, 28). Probes containing three substrate variants were

Selectivity determinants in class A sortases

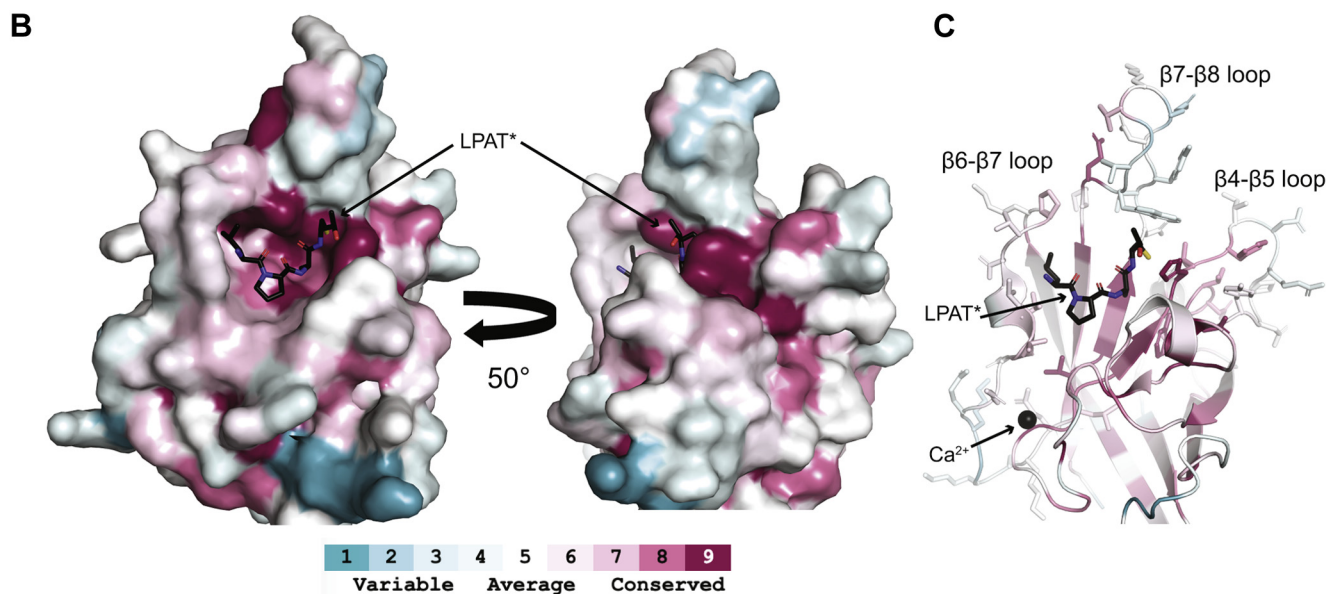
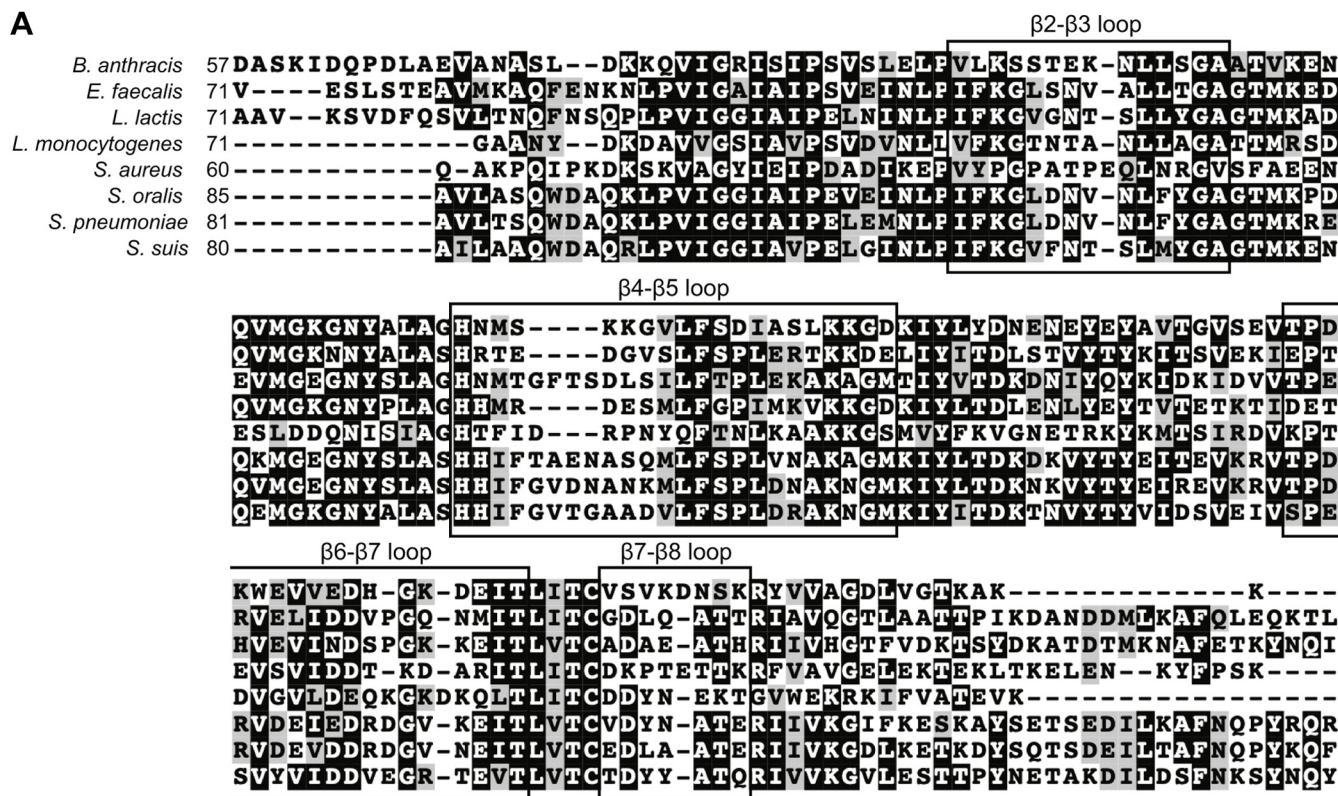


Figure 2. Sequence analysis and evolutionary conservation of class A sortases. *A*, multiple sequence alignment of previously studied class A sortase sequences (22). Alignment was performed using T-Coffee, and the figure was prepared using BoxShade. *Black boxes* indicate loop regions in saSrtA, as labeled. The *black box* for the β7-β8 loop cuts off the final five residues of the saSrtA loop (GVWEK) because the alignment is not correct in this region. *B*, the ConSurf server was used to investigate evolutionary conservation, using saSrtA (PDB ID: 2KID) as a template and a multiple sequence alignment of 400 class A sortase sequences. The results are shown on 2KID in surface representation (key below) and the LPAT* peptidomimetic is in *black sticks* and colored by heteroatom. Conservation is highest in the peptide-binding cleft and on the backside of the active site, where lipid II is hypothesized to bind. *C*, cartoon representation of the ConSurf results as described in panel *B*, with the side chain *sticks* of the β4-β5, β6-β7, and β7-β8 loops shown. SaSrtA binds calcium, which is shown as a *black sphere* and labeled. saSrtA, sortase A from *Staphylococcus aureus*.

initially prepared (Abz-LPATAG-K(Dnp), Abz-LPATGG-K(Dnp), Abz-LPATSG-K(Dnp), varying only at P1' in bold) and used to test the relative activity of our WT and chimeric enzymes. For simplicity, we have hereafter omitted the Abz,

K(Dnp), and C-terminal glycine from peptide descriptions. For comparing enzyme activity, a standard 2-h reaction time was utilized and an excess of H₂NOH was included to resolve the acyl enzyme intermediates. For consistency, all reactions were

also conducted in the presence of Ca^{2+} , which is a required cofactor for saSrtA. A reaction end point (indicated by the increase in Abz fluorescence) for all enzyme/substrate pairings was then expressed relative to averaged benchmark reactions of WT saSrtA with the standard LPATG substrate (Figs. 3A and S2A). This benchmark reaction was consistently found to give ~84% conversion to the expected transacylation products when independently monitored *via* RP-HPLC (Fig. S2B).

Based on our previous results, we predicted that spSrtA would show activity for all three peptides, while saSrtA would be selective for LPATG (22). Consistent with this prediction, our results confirmed that spSrtA was equally capable of processing all three substrates, whereas saSrtA was restricted to LPATG (Fig. 3A). Our assay also revealed a marked reduction in spSrtA activity *versus* saSrtA, which was not captured in our previous study, likely because of the extended reaction time (24 h) used in that work (22). With respect to the chimeric enzymes, our results clearly showed that the sequence of the $\beta 7$ – $\beta 8$ loop was a major determinant of activity and specificity. Specifically, the saSrtA_{pneumoniae} protein was completely inactive while spSrtA_{aureus} functionally mimicked the narrow substrate preference of the WT saSrtA enzyme (Fig. 3A). This result is consistent with recently published data (21). To verify that our sortases were cleaving substrates at the expected site, reactions exhibiting a normalized fluorescence value of 0.2 or higher were independently

monitored by RP-HPLC and LC-MS, which confirmed cleavage between P1 and P1' (Figs. 3B and S2, B–F, Table S1). Notably, reactions for HPLC and LC-MS characterization were conducted in the presence and absence of Ca^{2+} , which demonstrated that this cofactor was not required for the activity of spSrtA and spSrtA_{aureus}.

Continuing on with the SpSrtA_{aureus} chimera, we next wanted to determine if the Trp residue derived from the saSrtA loop played a significant role in enzyme activity. In WT saSrtA, the W194 residue (using saSrtA numbering) is known to affect enzyme activity of saSrtA and has been shown to interact with the threonine of the LPXTG motif (Fig. S3) (19, 20). We therefore expressed and purified the corresponding “W194A” mutant of spSrtA_{aureus} and tested this variant with A-, G-, and S-containing peptides in our assay. This mutation in saSrtA has previously been characterized with respect to enzymatic activity but was not previously investigated with respect to possible effects on P1' selectivity (20). Indeed, our W194A spSrtA_{aureus} protein exhibited a 17% reduction in reaction progress for LPATG, while retaining its preference for Gly-containing peptides (Fig. 3A). However, W194A spSrtA_{aureus} also revealed activity for the A- and S-containing peptides suggesting that the Trp residue acts as a selectivity filter (Fig. 3A).

Variability in position P1' selectivity and transpeptidase activity in *S. pneumoniae* SrtA $\beta 7$ – $\beta 8$ variants

In addition to the profound shift in substrate scope observed for spSrtA_{aureus}, we were also intrigued that the overall reactivity of this chimera for LPATG was comparable with that seen with WT saSrtA. This stood in sharp contrast to the reaction of LPATG with WT spSrtA, where reaction progress was nearly two-thirds lower within the 2-h reaction time of our assay (Fig. 3A). Based on this, we wondered if similar gains in reactivity for substrates other than LPATG could be achieved by substituting in residues from additional SrtA proteins that we had evaluated previously (22). These chimeras included the $\beta 7$ – $\beta 8$ loop residues from *B. anthracis*, *E. faecalis* (spSrtA_{faecalis}), *L. lactis* (spSrtA_{lactis}), *L. monocytogenes* (spSrtA_{monocytogenes}), *S. oralis* (spSrtA_{oralis}), and *S. suis* (spSrtA_{suis}) (Fig. 4A) (22). To avoid confusion in the numbering of loops with variable lengths, we will hereafter refer to the N-terminal positions of each $\beta 7$ – $\beta 8$ loop by numbering with respect to the catalytic Cys ($\beta 7$ – $\beta 8^{+1}$, $\beta 7$ – $\beta 8^{+2}$, etc.) that precedes the loop, whereas the C-terminal loop residue will be numbered relative to the catalytic Arg ($\beta 7$ – $\beta 8^{-1}$) (Fig. 4A).

The six chimeric proteins were expressed and purified using the same protocol as spSrtA_{aureus}, and as described in Experimental procedures. The purity of all proteins was validated by SDS-PAGE, and SEC was consistent with the isolated proteins being predominantly monomeric (Fig. S1). With the new chimeras in hand, we conducted an initial evaluation of relative activity using the LPATG, LPATA, and LPATS substrates described above. While the majority of constructs exhibited significant reactivity across all three substrates,

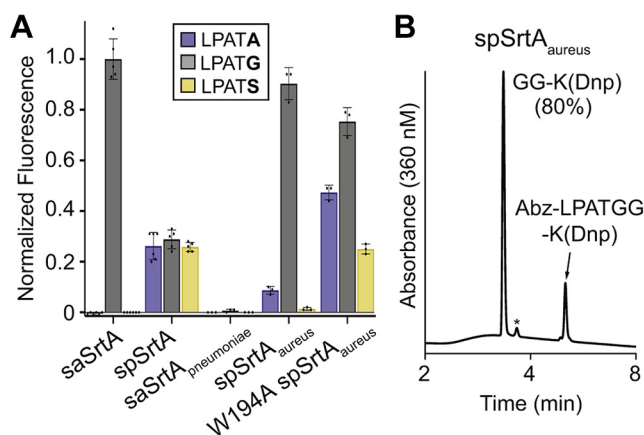


Figure 3. Interchanging $\beta 7$ – $\beta 8$ loops in class A sortases modulates substrate selectivity and activity for target sequences that vary at position P1' of the canonical LPXTG motif. A, comparison of substrate selectivity for WT saSrtA and spSrtA proteins, as well as $\beta 7$ – $\beta 8$ loop chimeras saSrtA_{pneumoniae}, spSrtA_{aureus}, and W194A spSrtA_{aureus}. Substrate cleavage was monitored *via* an increase in fluorescence at 420 nm from reactions of the fluorophore-quencher probes Abz-LPATGG-K(Dnp), Abz-LPATAG-K(Dnp), and Abz-LPATSG-K(Dnp) (represented as LPATG, LPATA, and LPATS) in the presence of excess hydroxylamine. Bar graphs represent the mean normalized fluorescence (\pm SD) from at least three independent experiments at the 2-h reaction timepoint, as compared with saSrtA and the peptide LPATG. B, representative HPLC chromatogram for the reaction of Abz-LPATGG-K(Dnp) and H_2NOH in the presence of spSrtA_{aureus}. This reaction was conducted in the presence of Ca^{2+} . Selective cleavage between the threonine (T) and glycine (G) residues was observed, with an overall conversion of 80% (* = Abz-LPAT-NHOH reaction product). Low peak intensity is due to the weak absorbance of Abz at 360 nm. Additional HPLC data for select substrate/enzyme pairings are provided in Figure S2, B–F. saSrtA, sortase A from *Staphylococcus aureus*; spSrtA_{monocytogenes}, $\beta 7$ – $\beta 8$ loop residues from *Listeria monocytogenes*; Abz, 2-aminobenzoyl fluorophore; Dnp, 2,4-dinitrophenyl quencher.

Selectivity determinants in class A sortases

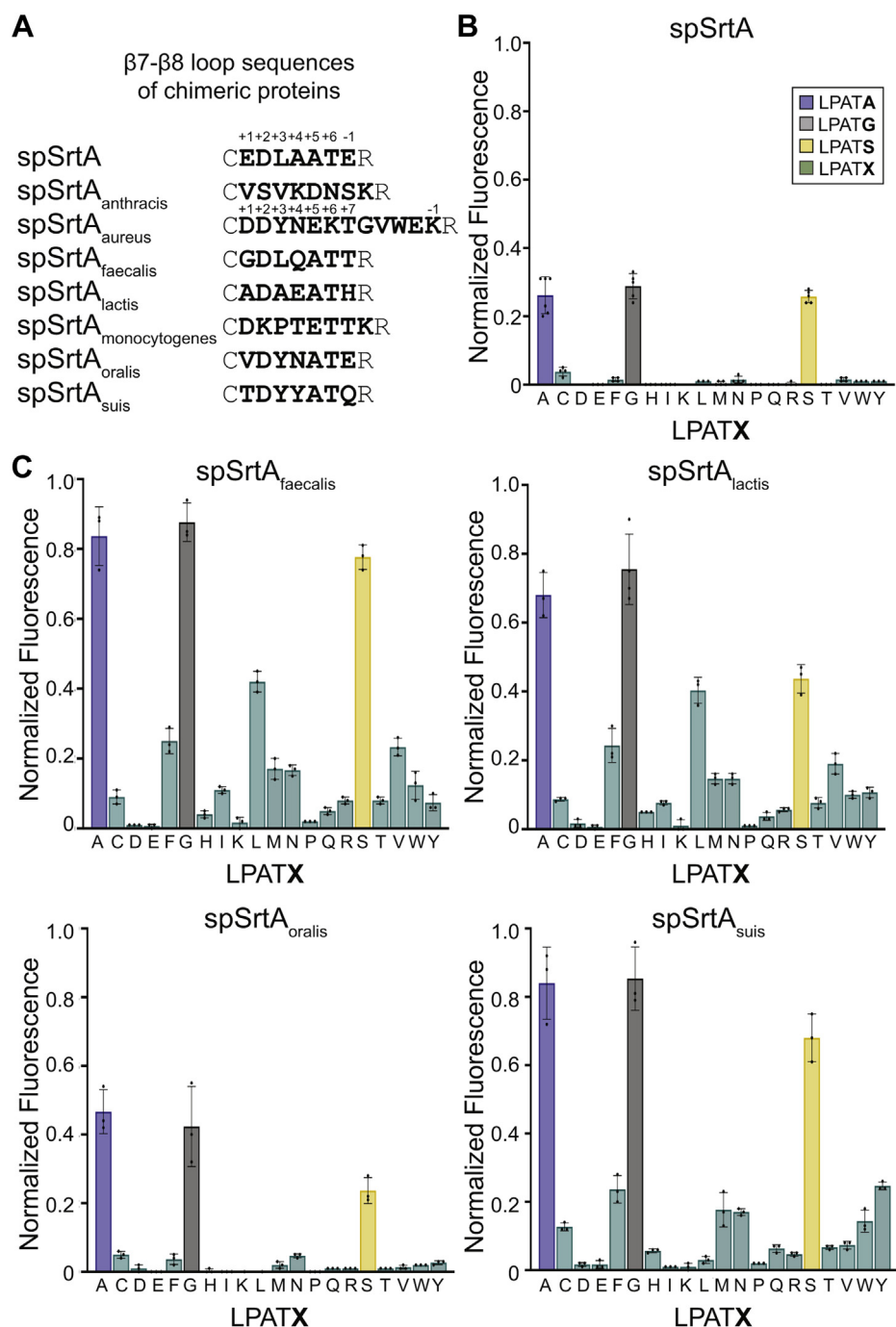


Figure 4. The sequence of the $\beta 7$ - $\beta 8$ loop dramatically affects selectivity and activity for spSrtA. A, the $\beta 7$ - $\beta 8$ loop sequences of the chimeric proteins used are listed, with representative numbering for residues in the $\beta 7$ - $\beta 8$ loop labeled for spSrtA and spSrtA_{aureus}. B and C, substrate selectivity profiles for WT spSrtA (B) and chimeric spSrtA variants (C). Substrate cleavage monitored via an increase in fluorescence at 420 nm from reactions of fluorophore-quencher probes with the generic structure Abz-LPATXG-K(Dnp) (LPATX) in the presence of hydroxylamine. Bar graphs represent the mean normalized fluorescence (\pm SD) from at least three independent experiments. spSrtA_{faecalis}, spSrtA with the $\beta 7$ - $\beta 8$ loop residues from *Enterococcus faecalis*; spSrtA_{lactis}, spSrtA with the $\beta 7$ - $\beta 8$ loop residues from *Lactococcus lactis*; spSrtA_{oralis}, spSrtA with the $\beta 7$ - $\beta 8$ loop residues from *Streptococcus oralis*; spSrtA_{suis}, spSrtA with the $\beta 7$ - $\beta 8$ loop residues from *Streptococcus suis*; Abz, 2-aminobenzoyl fluorophore; Dnp, 2,4-dinitrophenyl quencher.

spSrtA_{anthracis} and spSrtA_{monocytogenes}, containing the $\beta 7$ - $\beta 8$ loop residues from *B. anthracis* and *L. monocytogenes*, respectively, proved to be mostly inactive (Fig. S4A). For the remaining enzymes, the spSrtA_{oralis} protein behaved similarly to WT spSrtA, while spSrtA_{faecalis}, spSrtA_{lactis}, and spSrtA_{suis} showed improved performance for A-, G-, and S-containing substrates (Fig. 4, B and C and S4A). This was particularly

interesting in the case of spSrtA_{faecalis}, given that the WT SrtA enzyme from *E. faecalis* was previously shown to have poor reactivity for the same test substrates despite the use of higher enzyme loading and considerably longer reaction times (22).

Based on initial experiments with the A-, G-, and S-containing peptides, we next wanted to expand our peptide pool to assess the relative reactivity of our active chimeric spSrtA

variants for peptides containing all 20 amino acids at P1'. For comparison, a similar substrate profile was generated for WT spSrtA. As shown in Figure 4B, within the 2-h time frame of our assay, the WT protein was rather selective in its substrate recognition, with reactivity limited to A-, G-, and S-containing peptides. We note here that this somewhat limited substrate scope that appears to differ from the more promiscuous behavior reported previously for spSrtA. We attribute this to the fact that longer reaction times (24 h) and higher enzyme loadings (5-fold higher than the loading used here) were utilized in this earlier work (22). Similar to WT spSrtA, the spSrtA_{oralis} was limited to A-, G-, and S-containing peptides, albeit with slightly elevated reactivity in the case of LPATA and LPATG. Finally, we were intrigued to find that our spSrtA_{faecalis}, spSrtA_{lactis}, and spSrtA_{suis} proteins all show increased promiscuity for a variety of amino acids at P1' in our assay (Fig. 4C).

Overall, the spSrtA_{faecalis}, spSrtA_{lactis}, and spSrtA_{suis} proteins showed the largest increase in activity and promiscuity for this library of peptides. The spSrtA_{faecalis} and spSrtA_{lactis} proteins each recognized 15 of the 20 amino acids at P1' with normalized fluorescence values of ≥ 0.05 , while spSrtA_{suis} recognized 14 of the 20 (Fig. 4C). We chose 0.05 as a cut-off value to compare with the peptide activities of the spSrtA protein, which shows normalized fluorescence values of -0.02 to 0.02 for all non-G-, S-, or A-containing peptides, with the exception of LPATC (at 0.04 ± 0.01). Furthermore, spSrtA_{faecalis} and spSrtA_{suis} exhibited ~ 3 -fold higher reaction progress for the G-, S-, and A-containing peptides than spSrtA.

As verification of the results of our fluorescence assay, we also characterized a subset of enzyme/substrate combinations using RP-HPLC and LC-MS. Focusing on spSrtA_{faecalis}, we repeated reactions that exhibited normalized fluorescence values of ≥ 0.1 (LPATX, X = A, F, G, I, L, M, N, S, V, W, Y) (Fig. S4B). Reactions were conducted in the absence of Ca²⁺ to confirm that this cofactor was not required for activity. Successful substrate cleavage was observed in all cases, ranging from a high of 78% conversion in the case of LPATG, to only 6% conversion in the case of LPATW over 2 h at room temperature (RT) (Fig. S4B). Notably, the trends in relative substrate preferences observed by HPLC were consistent with those found in our original fluorescence assay (Fig. S5). In addition, while LC-MS characterization confirmed that substrate cleavage was occurring between the P1 and P1' of all sequences, certain substrates (LPATX, X = W, F, L, M, Y) containing bulky hydrophobic residues also produced alternate products arising from cleavage on the C-terminal side of P1' (Table S1, Fig. S4). In the case of LPATL, this alternate cleavage product was actually the major species obtained after reaction with spSrtA_{faecalis}. We note here that this capacity for alternate cleavage has been reported previously for WT spSrtA and thus appears to be maintained in the spSrtA_{faecalis} chimera (22).

Variability in position P1' selectivity and ligase activity in *S. pneumoniae* SrtA $\beta 7$ – $\beta 8$ variants

As a final assessment of the reactivity of the spSrtA_{faecalis} chimera, we next evaluated its ability to ligate amino acid

nucleophiles in place of the H₂NOH that was utilized in our fluorescence assay. For a series of test substrates (LPATX, X = A, S, V), spSrtA_{faecalis} was able to successfully ligate the corresponding free amino acid carboxamides (X-NH₂ = A-NH₂, S-NH₂, V-NH₂) with very good efficiency (Fig. 5). As expected from our fluorescence assay results, reaction progress with LPATV was slower than that observed for LPATA and LPATS. Specifically, reactions with LPATV required 8 h at RT to consume 85% of the initial peptide substrate, whereas reactions with LPATA/S exhibited $>95\%$ substrate conversion within 3 h. Importantly, the desired LPATX-NH₂ species was the major ligation product in all reactions as determined by LC-MS (Fig. 5, Table S1). Trace levels of substrate hydrolysis were also observed *via* LC-MS; however, the ratio of successful ligation to hydrolysis was 15:1 or better as estimated from mass spectral peak intensities. In reactions involving LPATV, we also detected low levels of substrate cleavage on the C-terminal side of the P1' valine residue. The extent of this alternate cleavage pathway was minimal, accounting for only

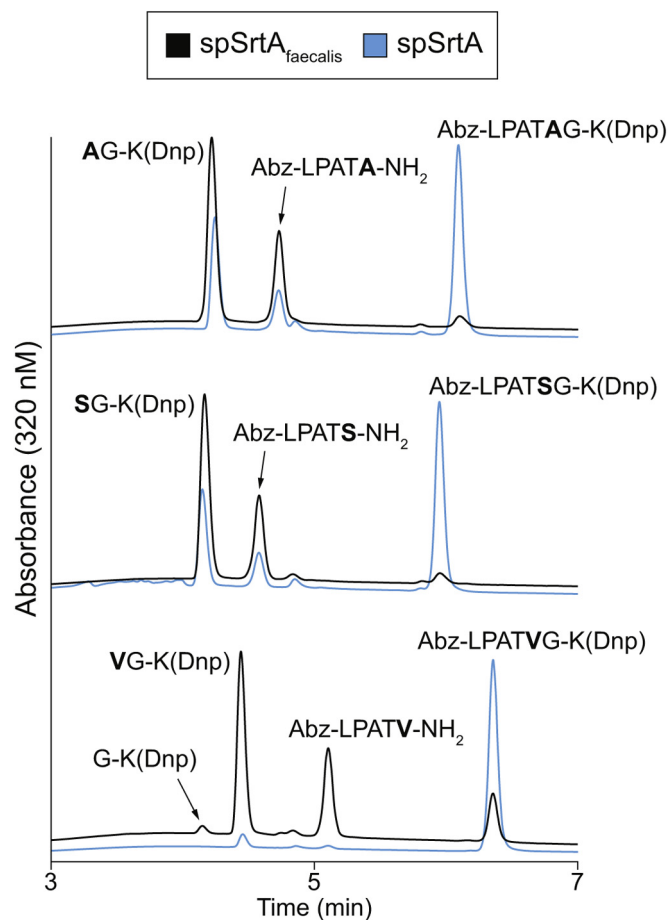


Figure 5. spSrtA_{faecalis} outperforms WT spSrtA in model amino acid ligation reactions. HPLC chromatograms (320 nm) for model ligations between Abz-LPATXG-K(Dnp) and excess X-NH₂ nucleophiles catalyzed by spSrtA_{faecalis} (black curves) or WT spSrtA (blue curves). Ligations were conducted in the absence of Ca²⁺. Chromatograms for LPATA/S represent the 3-h reaction timepoint, and chromatograms for LPATV correspond to the 8-h timepoint. All peak identities were confirmed *via* LC-MS (Table S1). spSrtA_{faecalis}, spSrtA with the $\beta 7$ – $\beta 8$ loop residues from *Enterococcus faecalis*; Abz, 2-aminobenzoyl fluorophore; Dnp, 2,4-dinitrophenyl quencher.

Selectivity determinants in class A sortases

~4% of the substrate cleavage events based on comparisons of HPLC peak areas for G-K(Dnp) and VG-K(Dnp) (Fig. 5). Interestingly, LC-MS characterization of these same reactions involving LPATV, V-NH₂, and spSrtA_{faecalis} failed to show clear evidence for the formation of ligation or hydrolysis products derived from the alternate cleavage pathway, potentially because of their low levels in solution.

For comparison, we also performed the same set of test ligations with WT spSrtA. In all cases, reaction progress was significantly reduced as compared with that of spSrtA_{faecalis} (Fig. 5). In particular, spSrtA exhibited minimal product formation with the LPATV system, representing a 10-fold reduction in reaction progress relative to spSrtA_{faecalis} for this atypical sortase substrate motif. Building from this result, an initial attempt to utilize the LPATV sequence as a handle for site-specific protein modification was made by installing this motif at the C-terminus of a full-size protein target. However, this protein substrate proved to be unreactive in the presence of both spSrtA_{faecalis} and WT spSrtA (data not shown).

Mutagenic investigation of the contribution of $\beta 7$ – $\beta 8$ loop residues

To dissect the contribution of each residue in the $\beta 7$ – $\beta 8$ loop of spSrtA, we made a series of alanine mutations. Specifically, we mutated all non-Ala residues to alanine to produce the following spSrtA variants: E208A, D209A, L210A, T213A, and E214A. Analytical SEC of the final protein preparations suggested that these mutants were predominantly monomeric (Fig. S6). When tested using our FRET-based activity assay, we did not see major effects on P1' selectivity in these mutants for G-, S-, and A-containing peptides; however, we did observe striking effects on overall reactivity (Fig. 6A). Specifically, a 2-fold increase in reaction progress was observed for the E208A and E214A mutations, suggesting that the native glutamic acid residues in spSrtA have a negative effect on its activity. The L210A and T213A mutations resulted in ~50% reduction in reaction conversion, and the D209A spSrtA protein was entirely unreactive (Fig. 6A).

Stereochemical basis of $\beta 7$ – $\beta 8$ variant selectivity and activity —(1) A stabilizing intraloop hydrogen bond

To gain a stereochemical understanding of our biochemical results, we analyzed available structures of class A sortases in the Protein Data Bank. To our knowledge, the 3D structure of an active, monomeric form of spSrtA has yet to be reported. Available crystal structures of the domain-swapped dimer show that the $\beta 7$ – $\beta 8$ loop is located at, and participates in, the dimer interface (PDB codes 4O8L, 4O8T, and 5DV0). Therefore, we chose to broaden our search to non-spSrtA structures, and in doing so, we identified three putative $\beta 7$ – $\beta 8$ loop-mediated interactions in class A sortases, which will be described in the following three sections.

First, we observed an intra-loop hydrogen bond in the $\beta 7$ – $\beta 8$ loops of several class A sortases (Fig. 6B). This Thr-mediated hydrogen bond is evident in SrtA proteins from

S. aureus (PDB code 2KID), *S. pyogenes* (3FN5), *Streptococcus mutans* (4TQX), and *L. monocytogenes* SrtA (ImSrtA) (5HU4), among others not shown (Fig. 6B). Notably, the *S. pyogenes* SrtA protein (3FN5) includes multiple rotamers in one protomer at the Thr position (T214), and the hydrogen bond with D210 is conserved in both (Fig. S7A). We predict that in spSrtA, this hydrogen bond will be formed between the side chains of residues $\beta 7$ – $\beta 8^{+2}$ D209 and $\beta 7$ – $\beta 8^{+6}$ T213. Indeed, we see this interaction in a homology model of the monomeric spSrtA protein generated using SWISS-MODEL (Fig. S7B).

For this model, the *S. pyogenes* SrtA structure (PDB code 3FN5) was used as a template because the crystallized form of this enzyme (*S. pyogenes* SrtA residues S81–T249) has 63% sequence identity with spSrtA (29–31). An alignment of our spSrtA model with 3FN5 revealed an overall RMSD of 0.083 Å over 567 main chain atoms. We further validated our model using structural alignments with a monomer extracted from the domain-swapped dimer structure (RMSD of 0.603 Å over 483 main chain atoms), as well as other SrtA structures from *Streptococcus* species, including those from *Streptococcus agalactiae* and *S. mutans* (PDB codes 3RCC (RMSD of 0.773 Å over 384 main chain atoms) and 4TQX (RMSD of 0.456 Å over 530 main chain atoms)), respectively (Fig. S7, C and D).

Our mutagenesis results are consistent with this proposed interaction. Specifically, the T213A mutation, which disrupts a potential intraloop hydrogen bond between the $\beta 7$ – $\beta 8^{+2}$ D209 and $\beta 7$ – $\beta 8^{+6}$ T213, reduced spSrtA activity by 54 to 73% for G-, S-, and A-containing peptides (Fig. 6A). When we attempted to reverse the hydrogen bond geometry with a D209T/T213D double mutant, the resulting enzyme exhibited only trace reactivity (Fig. 6A). Taken together, our findings are consistent with at least one stabilizing intraloop hydrogen bond being a generally conserved and functionally relevant feature of $\beta 7$ – $\beta 8$ SrtA loops. It should be noted, however, that the exact nature and location of this interaction likely varies. For example, in *Actinomyces oris* SrtA, this hydrogen bond is observed between $\beta 7$ – $\beta 8^{+3}$ D227 and $\beta 7$ – $\beta 8^{+11}$ S219 (Fig. S7E).

Stereochemical basis of $\beta 7$ – $\beta 8$ variant selectivity and activity —(2) A noncovalent interaction between the $\beta 6^{-2}$ residue and $\beta 7$ – $\beta 8$ loop

In addition to an intraloop hydrogen bond, we observe interactions between the $\beta 7$ – $\beta 8$ loop and $\beta 6^{-2}$ residues in multiple class A sortase structures (Fig. 6C). For example, K196 ($\beta 7$ – $\beta 8^{-1}$) of saSrtA interacts with the $\beta 6^{-2}$ D160 in several of the states of the reported NMR structure, PDB code: 2KID (Fig. 6C) (19). We also see a reasonable electrostatic interaction distance for the $\beta 7$ – $\beta 8^{-1}$ and $\beta 6^{-2}$ residues (K195 and E165, respectively) in several of the NMR states for baSrtA (PDB code 2RUI), as well as in the domain-swapped dimer structure of spSrtA (PDB code: 4O8L), which shows the E214 $\beta 7$ – $\beta 8$ loop residue of one protomer interacting with R184 (Fig. 6C). Interestingly, the nature of this interaction can change, as in the case of *Actinomyces oris* SrtA, where both residues are hydrophobic leucine residues (L186 and L228 for the $\beta 6^{-2}$ and $\beta 7$ – $\beta 8^{-1}$ residues, respectively) (Fig. 6C).

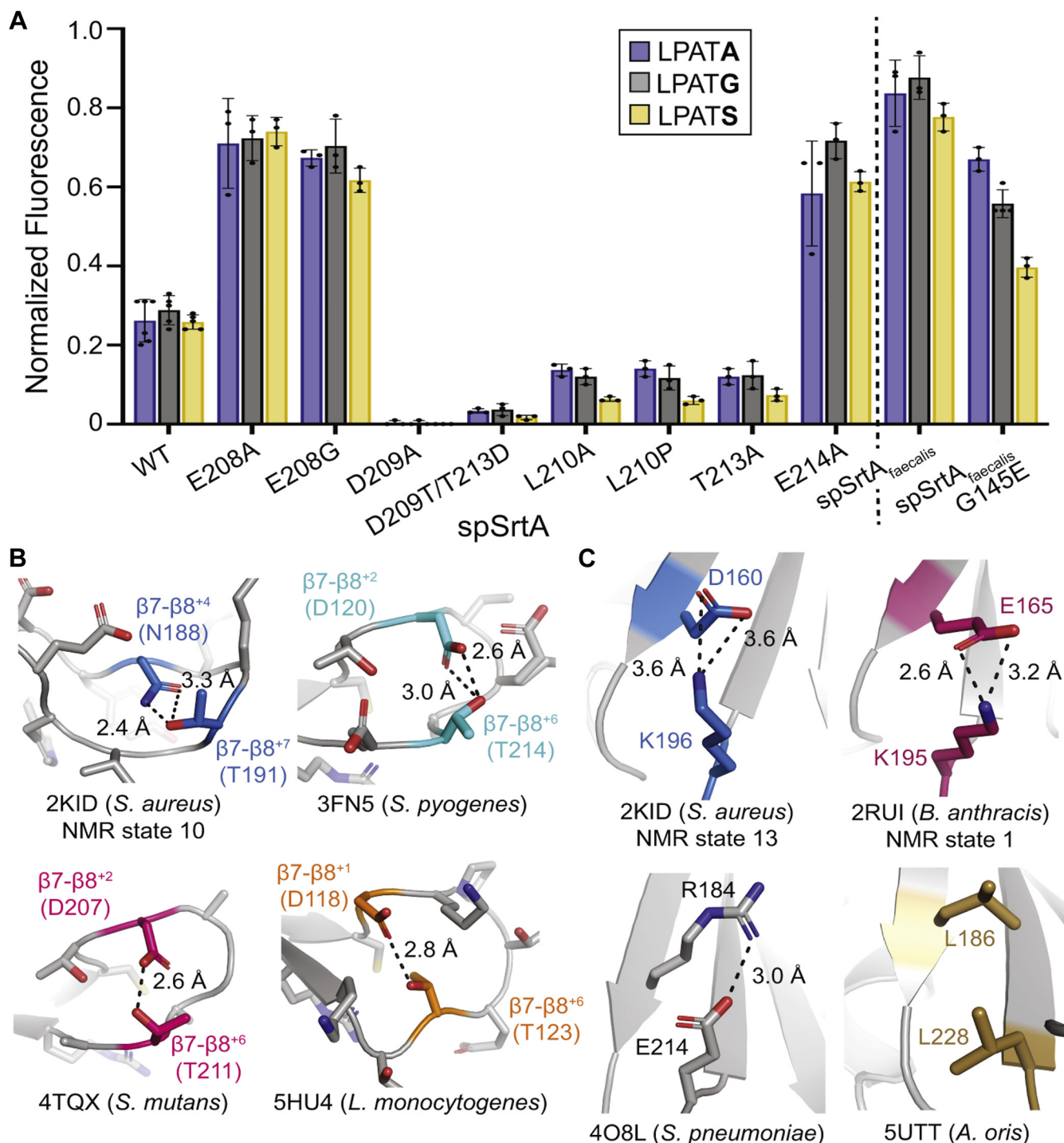


Figure 6. Residues in the $\beta 7$ - $\beta 8$ loop participate in interactions that affect enzyme activity and selectivity, based on structural analyses and mutagenesis. *A*, enzyme assays on mutant spSrtA and spSrtA_{faecalis} variants (fluorophore-quencher substrate assay conditions identical to those described in the legend to Fig. 3). Bar graphs represent the mean normalized fluorescence values (\pm SD) from at least three independent experiments. *B* and *C*, all SrtA structures are in gray ribbons. The $\beta 7$ - $\beta 8$ loop side chains are all in stick representation and colored by heteroatom. Residues that participate in hydrogen bonds have non-gray carbons and are labeled. The hydrogen bonds are shown with a black dashed line, with measurements indicated. For NMR structures (PDB IDs 2KID and 2RUI), the state that was used for the image is labeled. Although not all NMR states contained the interaction indicated, in both cases, there were several states that revealed measurements consistent with a noncovalent interaction. Any side-chain sticks are colored by heteroatom (O = red, N = blue). *B*, $\beta 7$ - $\beta 8$ intraloop hydrogen bond. The structures in this figure are as follows: *S. aureus* SrtA (2KID, blue carbons), *S. pyogenes* SrtA (3FN5, cyan carbons), *S. mutans* SrtA (4TQX, pink carbons), and *L. monocytogenes* SrtA (5HU4, orange carbons). *C*, interaction between the $\beta 7$ - $\beta 8$ loop and $\beta 6$ strand. The structures in this figure are as follows: *S. aureus* SrtA (PDB ID 2KID, blue carbons), *B. anthracis* SrtA (2RUI, dark pink carbons), *S. pneumoniae* SrtA (4O8L, gray carbons), and *A. oris* SrtA (5UTT, gold carbons). spSrtA_{faecalis}, spSrtA with the $\beta 7$ - $\beta 8$ loop residues from *Enterococcus faecalis*.

Selectivity determinants in class A sortases

Our results suggest that this $\beta 7$ – $\beta 8$ loop/ $\beta 6^{-2}$ interaction has a negative effect on sortase activity. This is supported by the >2 -fold increase in substrate conversion for both the $\beta 7$ – $\beta 8^{-1}$ E214A and $\beta 7$ – $\beta 8^{+1}$ E208A spSrtA mutants as compared with WT spSrtA (Fig. 6A). While the reported domain-swapped dimer structure of spSrtA does not exhibit an obvious interaction between $\beta 7$ – $\beta 8^{+1}$ (E208) and $\beta 6^{-2}$ (R184), we note our spSrtA homology model does suggest this interaction is likely, with a distance between a guanidinium nitrogen atom of R184 and a side-chain carboxylate oxygen atom on E208 equal to 2.7 Å (Fig. S7B). As further evidence, given that this $\beta 7$ – $\beta 8^{+1}$ position is a glycine in the more active spSrtA_{faecalis} chimera, we expressed, purified, and tested the activity of two contrasting mutants at this site: E208G spSrtA (which would eliminate the putative $\beta 7$ – $\beta 8^{+1}$ / $\beta 6^{-2}$ interaction) and G145E spSrtA_{faecalis} (*E. faecalis* SrtA numbering, which would restore the putative $\beta 7$ – $\beta 8^{+1}$ – $\beta 6^{-2}$ interaction). Our results are consistent with our predictions, and we saw a 21 to 49% reduction in substrate conversion for the reaction of G-, S-, and A-containing peptides with G145E spSrtA_{faecalis} as compared with the initial spSrtA_{faecalis} chimera (Fig. 6A). In contrast, a >2 -fold increase in reaction progress was observed for E208G spSrtA relative to WT spSrtA (Fig. 6A).

Overall, our data are consistent with recent work where a triple mutant of *S. pyogenes* SrtA (E189H/V206I/E215A, where E215A is a mutation at the $\beta 7$ – $\beta 8^{-1}$ position) resulted in 6.6-fold enhanced catalytic efficiency (32). In addition, K196T in the catalytically enhanced pentamutant saSrtA protein is also located at the $\beta 7$ – $\beta 8^{-1}$ position (8). Taken together, the data support that a noncovalent interaction between the $\beta 7$ – $\beta 8$ loop and $\beta 6$ strand negatively affects SrtA activity.

Stereochemical basis of $\beta 7$ – $\beta 8$ variant selectivity and activity —(3) An interaction between the $\beta 7$ – $\beta 8$ and $\beta 4$ – $\beta 5$ loops

NMR structures of saSrtA and baSrtA proteins in the unbound and bound states (PDB codes: 1IJA, 2KID, 2KW8, and 2RUI) suggest distinct mechanisms of substrate binding (Fig. S8). In saSrtA, the $\beta 7$ – $\beta 8$ loop is ordered in both states but moves upon ligand binding (Fig. S8A). In contrast, a unique N-terminal appendage in baSrtA regulates active site accessibility, as previously described, and the $\beta 7$ – $\beta 8$ loop transitions from a disordered-to-ordered state upon substrate binding (Fig. S8B) (33). In both proteins, however, binding of ligand corresponds to a shift in the position of the $\beta 7$ – $\beta 8$ loop such that it is located much closer to the $\beta 4$ – $\beta 5$ loop (Fig. 7).

In analyses of the previously published baSrtA structures, the authors of this work mention that the $\beta 4$ – $\beta 5^{+2}$ / $\beta 4$ – $\beta 5^{+3}$ positions (for baSrtA, these are $\beta 4$ – $\beta 5^{+2}$ M128 and $\beta 4$ – $\beta 5^{+3}$ S129) play a role in stabilizing a hydrophobic residue, the $\beta 7$ – $\beta 8^{+3}$ V190, upon ligand binding (Fig. 7) (33). The corresponding residues in saSrtA are F64, I65, and Y129, respectively, which would also enable a favorable hydrophobic interaction (Fig. 7). The lmSrtA structure (PDB ID: 5HU4) also shows a similar interaction: $\beta 4$ – $\beta 5^{+2}$ M129 and $\beta 4$ – $\beta 5^{+3}$ R130 with P191 in the $\beta 7$ – $\beta 8^{+3}$ position (note: M129 is the full-length protein numbering according to UniProt ID

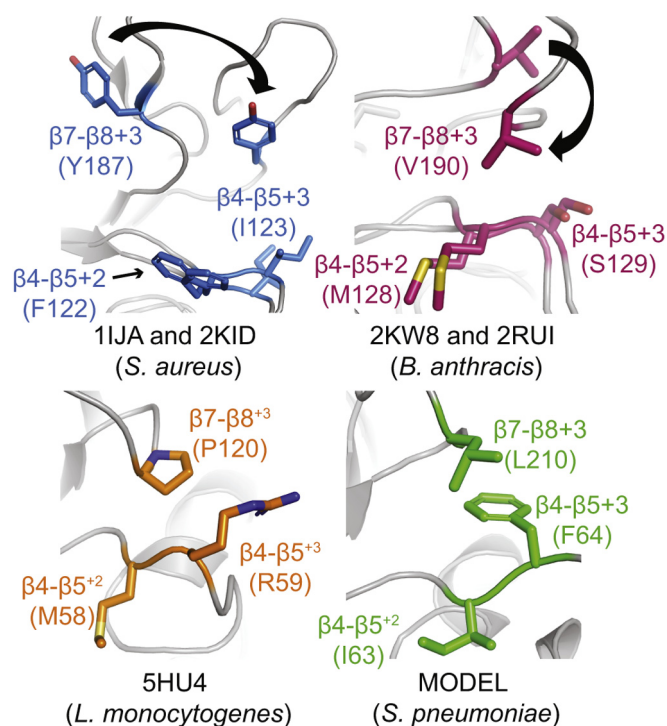


Figure 7. Interaction between the $\beta 7$ – $\beta 8$ and $\beta 4$ – $\beta 5$ loops. All SrtA structures are in gray ribbons. The $\beta 7$ – $\beta 8$ loop side chains are all in stick representation and colored by heteroatom. The structures in this figure are as follows: *S. aureus* (unbound: 1IJA and bound: 2KID, blue carbons), *B. anthracis* (unbound: 2KW8 and bound: 2RUI, dark pink carbons), *L. monocytogenes* (5HU4, orange carbons), and *Streptococcus pneumoniae* (homology model (Fig. S7), green carbons). The arrows indicate residue movement from the unbound to bound structures.

SRTA_LISMO, but is M58 in 5HU4, R130 is R59, and P191 is P120), in the unbound state (Fig. 7). Finally, our spSrtA model suggests a potential interaction between the $\beta 4$ – $\beta 5^{+2}$ I63 and $\beta 4$ – $\beta 5^{+3}$ F64 residues with $\beta 7$ – $\beta 8^{+3}$ L210 (Fig. 7).

In agreement with this proposed interaction between the $\beta 7$ – $\beta 8$ and $\beta 4$ – $\beta 5$ loops, we observed that our L210A spSrtA mutant reduced the activity of the protein by 46%, 59%, and 77% for the A-, G-, and S-containing peptides, respectively (Fig. 6A). While the alanine methyl side chain retains the hydrophobic character of the WT leucine, we speculate that its reduced size is insufficient to space the gap between the $\beta 7$ – $\beta 8$ and $\beta 4$ – $\beta 5$ loops, and thus this critical interaction is disrupted. Along these lines, we expressed, purified, and tested an L210P spSrtA mutant, which substitutes L210 with the proline residue found in spSrtA_{monocytogenes}, a chimeric protein which was essentially inactive in our hands (Fig. S4A). In our substrate assay, this mutant produced results identical to those of L210A, further supporting the importance of a specific interaction between the $\beta 7$ – $\beta 8^{+3}$ residue and $\beta 4$ – $\beta 5$ loop for sortase activity (Fig. 6A).

Discussion

Although target sequence recognition by *S. aureus* SrtA is rigidly selective for a P1' glycine, this is not true of all class A sortases, such as those from *S. pneumoniae* and *S. pyogenes* (10, 14, 29, 34). Building from our previous work,

in which spSrtA was found to accept peptides containing Gly, Ala, and Ser and other residues at P1', we have shown here that this broadened substrate scope can be attributed to the sequence of the $\beta 7$ – $\beta 8$ loop (26). Moreover, variations in $\beta 7$ – $\beta 8$ loop sequences can substantially impact the overall enzyme activity, affording chimeric sortases that outperform their WT counterpart *in vitro*. Together with others, the present study implicates all of the variable loops in class A sortases as being important determinants of enzyme function (12, 18, 21, 33).

With respect to structure, we propose three interactions that are facilitated by residues in the $\beta 7$ – $\beta 8$ loop of spSrtA. They are as follows: (1) an intraloop hydrogen bond that positively affects catalytic efficiency, typically mediated by a threonine residue at the $\beta 7$ – $\beta 8^{+6}$ or $\beta 7$ – $\beta 8^{+7}$ position, (2) an interaction that hinders enzyme activity between the $\beta 7$ – $\beta 8$ loop and $\beta 6^{-2}$ residues, and (3) a positive interaction between the $\beta 7$ – $\beta 8^{+3}$ and $\beta 4$ – $\beta 5^{+2}/\beta 4$ – $\beta 5^{+3}$ residues, typically of hydrophobic nature. Notably, there appear to be other features in this structurally conserved loop that are unique to certain class A sortases. These include the W194 residue of saSrtA, which specifically interacts with the P1 position of the CWSS and likely acts as a selectivity filter, based on our W194A spSrtA_{aureus} data (Fig. 3A) (19). Others identified a disordered-to-ordered transition of the baSrtA $\beta 7$ – $\beta 8$ loop, as well as regulation by an N-terminal appendage, although more research is needed to determine whether or not this is shared by other class A sortases (33). Furthermore, while the spSrtA $\beta 7$ – $\beta 8$ loop is seven residues in length, several class A sortase loops, for example, those of saSrtA, baSrtA, and ImSrtA studied here, are longer than seven residues. Stabilizing interactions mediated by backbone atoms likely vary in loops of differing lengths, a characteristic that was not studied in detail here. Future work is also needed to assess how the loop length influences the identified interactions described here in the context of sortase A enzymes, as well as other position-specific interactions found in other sortase classes.

In addition to informing our fundamental understanding of sortase substrate recognition, this work also has implications for the further development of SML as a protein engineering tool (3, 35). Through exchange of $\beta 7$ – $\beta 8$ loop residues between class A sortases, we have generated chimeras such as spSrtA_{faecalis} and spSrtA_{lactis}, with measurable activity against peptides possessing 15 of the 20 amino acids at P1'. Notably, stability or minor changes in the folded structures between variants were not determined and these could manifest as some of the differences in enzyme function observed. With additional development, each of these sortase chimera/substrate combinations potentially offers a new handle for *in vitro* SML applications. While preliminary attempts here to modify a protein target displaying an LPATV sequence using spSrtA_{faecalis} were unsuccessful, we consider it likely that optimization of the placement of the LPATV site may restore reactivity. This includes examination of the accessibility requirements for the LPATV sequence and assessment of the impact of residues N- or C-terminal to the core LPATV motif. Similar factors are known to affect the

success of SML reactions with the widely used saSrtA/LPXTG system (36–38) and may need to be evaluated for our chimeras.

If successful, the development of these new sortase/substrate pairs has exciting consequences for SML engineering efforts: (1) it increases options for dual-labeling single proteins or multiplexed labeling of multiple proteins in the same systems (11, 39, 40), and (2) it may reduce the need to mutate naturally occurring protein sequences to render their termini compatible with SML. For example, using our previously published program, *MotifAnalyzer*, we found 190 instances of LPXTG in 189 unique proteins in the human proteome. However, if the P1' position is now flexible, this number becomes 3606 instances of LPXTX in 2930 unique proteins (41).

Finally, the three variable loops mentioned here ($\beta 4$ – $\beta 5$, $\beta 6$ – $\beta 7$, and $\beta 7$ – $\beta 8$) are conserved in all classes of sortases (Fig. 1B), and previous work determining and engineering sortase selectivity of different classes, for example, sortase B, suggests similar roles for these loops in substrate recognition (5, 18, 42). Developing a deeper understanding of how residues in these loops affect substrate selectivity in all sortase classes may enable dramatic expansion of the sortase “toolbox” (Fig. 8), potentially allowing the development of ligases that are tailored to the needs of specific protein targets while also limiting off-target effects (5, 11–14, 17, 22). In the over 20 years since saSrtA was discovered, the sortase superfamily has proven to be both a workhorse for protein engineering efforts and an exciting system for future discoveries and insight into the stereochemistry and mechanisms of target recognition.

Experimental procedures

Protein expression and purification

WT spSrtA and saSrtA proteins were expressed and purified as previously described (22). All other constructs, including chimeric and mutant proteins, were purchased from GenScript in the pET28a(+) vector. In general, protein expression and purification protocols were very similar to those previously described (22). Briefly, plasmids were transformed into *E. coli*

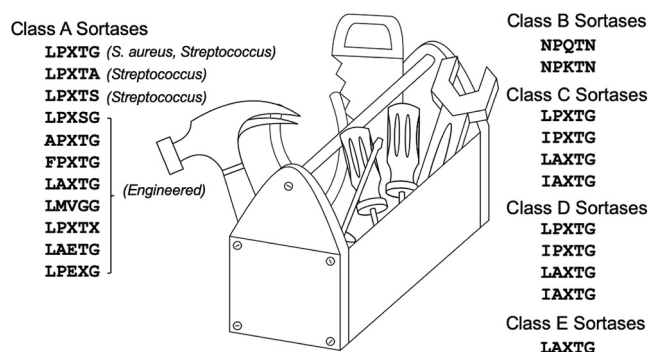


Figure 8. Building a sortase toolbox for SML experiments. Work from ourselves and others can be used to create a sortase “toolbox” for SML experiments, taking advantage of the various sequence motifs, both endogenous and engineered. Recognition sequences for various sortase subclasses are described in (5, 11–14, 17, 22). SML, sortase-mediated ligation.

Selectivity determinants in class A sortases

BL21 (DE3) competent cells and grown in LB media, with protein induction at A_{600} 0.6 to 0.8 using 0.15 M IPTG for 18 to 20 h at 18 °C.

After cell harvest in the lysis buffer [0.05 M Tris, pH 7.5, 0.15 M NaCl, and 0.0005 M EDTA], the protein was purified using a 5-ml HisTrap HP column (GE Life Sciences, now Cytiva), using wash [0.05 M Tris, pH 7.5, 0.15 M NaCl, 0.02 M imidazole, pH 7.5, and 0.001 M TCEP] and elution [wash buffer, with 0.3 M imidazole, pH 7.5] buffers. SEC was conducted using a HiLoad 16/600 Superdex 75 column (GE Life Sciences, now Cytiva) in the SEC running buffer (0.05 M Tris, pH 7.5, 0.15 M NaCl, and 0.001 M TCEP). Purified protein corresponding to the monomeric peak was concentrated using an Amicon Ultra-15 Centrifugal Filter Unit (10,000 NWML) and analyzed by SDS-PAGE and analytical SEC (Figs. S1 and S6). Protein not immediately used was flash-frozen in the SEC running buffer and stored at -80 °C.

Peptide synthesis

Detailed synthetic procedures are provided in the [supporting information](#). Briefly, all peptides were synthesized *via* manual Fmoc solid-phase peptide synthesis. Peptides were synthesized either individually or in tandem using Fmoc Rink amide MBHA resin or SynPhase lantern solid supports. All other materials, including suitably protected Fmoc amino acids, and reagents for coupling, deprotection, and resin cleavage were obtained from commercial sources and used without further purification. All peptides were purified using RP-HPLC, and their identities were confirmed *via* ESI-MS. Before use in sortase-catalyzed transacylation reactions, each purified peptide was prepared as a concentrated stock solution in DMSO and/or water (see [supporting information](#) for details).

Fluorescence assay for sortase activity

Reactions were performed in a Costar round-bottom, black, 96-well plate at a 100- μ l reaction volume under the following conditions: 5 μ M sortase, 50 μ M peptide substrate, and 5 mM hydroxylamine nucleophile. All reactions contained 10% (*v/v*) 10 \times sortase reaction buffer (500 mM Tris, pH 7.5, 1500 mM NaCl, and 100 mM CaCl_2). Reactions also contained residual DMSO from the peptide stock solutions (0.5–1.5% (*v/v*), with the exception of the Phe- and Val-containing peptides at 5%). The peptides containing phenylalanine or valine required 5% (*v/v*) DMSO for solubility under the reaction conditions. 1 mM TCEP was also included in reactions utilizing the Abz-LPATCG-K(Dnp) substrate. Reactions were initiated by the addition of the sortase enzyme, which were prepared as 10 \times stock solutions in 50 mM Tris, pH 7.5, 150 mM NaCl, and 1 mM TCEP. Microplates were analyzed using a BioTek Synergy H1 plate reader. The fluorescence intensity of each well was measured at 2-min time intervals over a 2-h period at RT ($\lambda_{\text{ex}} = 320$ nm, $\lambda_{\text{em}} = 420$ nm, and detector gain = 75). All reactions were performed at least in triplicate, and all of the raw fluorescence data utilized in this study are provided in [Table S2](#). For each substrate sequence, the background

fluorescence of the intact peptide in the absence of enzyme was subtracted from the observed experimental data. Background-corrected fluorescence data were then normalized to the fluorescence intensity of a benchmark reaction between WT saSrtA and Abz-LPATGG-K(Dnp) (Fig. S2A).

HPLC and LC-MS characterization of sortase-catalyzed reactions

Select pairings of sortase enzyme (5 μ M or 10 μ M for the X-NH₂ reactions), substrate (50 μ M), and nucleophile (5 mM H₂NOH or X-NH₂) were repeated in the presence or absence of Ca²⁺ under reaction conditions that were otherwise identical to those described above for the fluorescence assay. These reactions were then analyzed using a Dionex UltiMate 3000 HPLC system interfaced with an Advion CMS expression mass spectrometer. Separations were achieved with a Phenomenex Kinetex 2.6 μ M C18 100 Å column (100 \times 2.1 mm) (aqueous [95% water, 5% MeCN, 0.1% formic acid]/MeCN [0.1% formic acid] mobile phase at 0.3 ml/min, method: hold 10% MeCN for 0.0–0.5 min, linear gradient of 10–90% MeCN for 0.5–7.0 min, hold 90% MeCN for 7.0–8.0 min, linear gradient of 90–10% MeCN for 8.0–8.1 min, re-equilibrate at 10% MeCN for 8.1–13.25 min)).

Sequence and structural analyses

All sequences were downloaded from either the NCBI database or UniProt, as indicated (43–45). Sequence alignments were performed using MAFFT, T-Coffee, or BlastP (23, 46, 47). Visualization of our T-Coffee alignment was done using BoxShade. ConSurf analyses were performed using the online server, with 2KID as a template and insertion or our own multiple sequence alignment performed using MAFFT (24, 25). Alignments were visualized using Jalview (48). Homology modeling was performed using the SWISS-MODEL web interface (30, 31). Structural analyses and figure rendering were done using PyMOL. Enzyme assay graphs were prepared using GraphPad Prism 9.1.2.

Data availability

All data are contained in the article and the [supporting information](#).

Supporting information—This article contains [supporting information](#) (22, 49, 50).

Acknowledgments—The authors would like to thank the other members of the Amacher and Antos laboratories for helpful discussions and assistance. Start-up funds from Western Washington University also contributed to this project.

Author contributions—I. M. P., S. A. S., J. D. V., K. L. H., J. M. A., and J. F. A. data curation; I. M. P., S. A. S., J. D. V., K. L. H., and J. M. A. formal analysis; I. M. P., J. D. V., and J. M. A. validation; I. M. P., S. A. S., J. D. V., D. A. J., M. G., K. J., J. E. S., H. M. K., K. L. H., J. M. A., and J. F. A. investigation; I. M. P., J. D. V., J. M. A., and J. F. A. methodology; I. M. P., S. A. S., J. D. V., D. A. J., M. G., K. J., J. E. S., H. M. K., K. L. H., J. M. A., and J. F. A. writing—review

and editing; J. D. V., K. L. H., J. M. A., and J. F. A. visualization; K. L. H., J. M. A., and J. F. A. funding acquisition; J. M. A. and J. F. A. conceptualization; J. M. A. and J. F. A. resources; J. M. A. and J. F. A. supervision; J. M. A. and J. F. A. writing—original draft.

Funding and additional information—J. F. A. was funded by NSF CHE-1904711 and J. M. A. was funded by a Cottrell Scholar Award from the Research Corporation for Science Advancement. K. L. H. was supported by the NIAID, National Institutes of Health, Award Number F32AI145111. In addition, I. M. P. received an Elwha Undergraduate Summer Research Award and D. A. J. received a Joseph and Karen Morse Student Research in Chemistry Fellowship to fund summer research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interests—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: Abz, 2-aminobenzoyl fluorophore; baSrtA, *Bacillus anthracis* SrtA; CWSS, cell wall sorting signal; Dnp, 2,4-dinitrophenyl quencher; ImSrtA, *Listeria monocytogenes* SrtA; saSrtA, sortase A from *Staphylococcus aureus*; SEC, size-exclusion chromatography; SML, sortase-mediated ligation; spSrtA, *Streptococcus pneumoniae* SrtA; spSrtA_{faecalis}, β 7– β 8 loop residues from *Enterococcus faecalis*; spSrtA_{lactis}, β 7– β 8 loop residues from *Lactococcus lactis*; spSrtA_{monocytogenes}, β 7– β 8 loop residues from *Listeria monocytogenes*; spSrtA_{oralis}, β 7– β 8 loop residues from *Streptococcus oralis*; spSrtA_{suis}, β 7– β 8 loop residues from *Streptococcus suis*.

References

- Spirig, T., Weiner, E. M., and Clubb, R. T. (2011) Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* **82**, 1044–1059
- Jacobitz, A. W., Kattke, M. D., Wereszczynski, J., and Clubb, R. T. (2017) Sortase transpeptidases: Structural biology and catalytic mechanism. *Adv. Protein Chem. Struct. Biol.* **109**, 223–264
- Antos, J. M., Truttmann, M. C., and Ploegh, H. L. (2016) Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* **38**, 111–118
- Di Girolamo, S., Puorger, C., Castiglione, M., Vogel, M., Gébleux, R., Briendl, M., Hell, T., Beerli, R. R., Grawunder, U., and Lipps, G. (2019) Characterization of the housekeeping sortase from the human pathogen propionibacterium acnes: First investigation of a class F sortase. *Biochem. J.* **476**, 665–682
- Bradshaw, W. J., Davies, A. H., Chambers, C. J., Roberts, A. K., Shone, C. C., and Acharya, K. R. (2015) Molecular features of the sortase enzyme family. *FEBS J.* **282**, 2097–2114
- Marraffini, L. A., DeDent, A. C., and Schneewind, O. (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 192–221
- Mazmanian, S. K., Liu, G., Ton-That, H., and Schneewind, O. (1999) *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760–763
- Chen, L., Dorr, B. M., and Liu, D. R. (2011) A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11399–11404
- Freund, C., and Schwarzer, D. (2021) Engineered sortases in peptide and protein chemistry. *ChemBiochem* **22**, 1347–1356
- Kruger, R. G., Otvos, B., Frankel, B. A., Bentley, M., Dostal, P., and McCafferty, D. G. (2004) Analysis of the substrate specificity of the *Staphylococcus aureus* sortase transpeptidase SrtA. *Biochemistry* **43**, 1541–1551
- Dorr, B. M., Ham, H. O., An, C., Chaikof, E. L., and Liu, D. R. (2014) Reprogramming the specificity of sortase enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13343–13348
- Piotukh, K., Geltinger, B., Heinrich, N., Gerth, F., Beyermann, M., Freund, C., and Schwarzer, D. (2011) Directed evolution of sortase A mutants with altered substrate selectivity profiles. *J. Am. Chem. Soc.* **133**, 17536–17539
- Podracky, C. J., An, C., DeSousa, A., Dorr, B. M., Walsh, D. M., and Liu, D. R. (2021) Laboratory evolution of a sortase enzyme that modifies amyloid- β protein. *Nat. Chem. Biol.* **17**, 317–325
- Antos, J. M., Chew, G.-L., Guimaraes, C. P., Yoder, N. C., Grotenbreg, G. M., Popp, M. W.-L., and Ploegh, H. L. (2009) Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc.* **131**, 10800–10801
- Clancy, K. W., Melvin, J. A., and McCafferty, D. G. (2010) Sortase transpeptidases: Insights into mechanism, substrate specificity, and inhibition. *Biopolymers* **94**, 385–396
- Bentley, M. L., Lamb, E. C., and McCafferty, D. G. (2008) Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from *Staphylococcus aureus*. *J. Biol. Chem.* **283**, 14762–14771
- Schmohl, L., Bierlmeier, J., Gerth, F., Freund, C., and Schwarzer, D. (2017) Engineering sortase A by screening a second-generation library using phage display. *J. Pept. Sci.* **23**, 631–635
- Bentley, M. L., Gaweska, H., Kielec, J. M., and McCafferty, D. G. (2007) Engineering the substrate specificity of *Staphylococcus aureus* Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. *J. Biol. Chem.* **282**, 6571–6581
- Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E., and Clubb, R. T. (2009) The structure of the *Staphylococcus aureus* sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized. *J. Biol. Chem.* **284**, 24465–24477
- Ton-That, H., Mazmanian, S. K., Alksne, L., and Schneewind, O. (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis. *J. Biol. Chem.* **277**, 7447–7452
- Wójcik, M., Szala, K., van Merkerk, R., Quax, W. J., and Boersma, Y. L. (2020) Engineering the specificity of *Streptococcus pyogenes* sortase A by loop grafting. *Proteins* **88**, 1394–1400
- Nikghalb, K. D., Horvath, N. M., Prelesnik, J. L., Banks, O. G. B., Filipov, P. A., Row, R. D., Roark, T. J., and Antos, J. M. (2018) Expanding the scope of sortase-mediated ligations by using sortase homologues. *ChemBiochem* **19**, 185–195
- Katoh, K., Rozewicki, J., and Yamada, K. D. (2019) MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016) ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350
- Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., and Ben-Tal, N. (2005) ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302
- Perry, A. M., Ton-That, H., Mazmanian, S. K., and Schneewind, O. (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an *in vivo* peptidoglycan substrate for sortase-catalyzed surface protein anchoring. *J. Biol. Chem.* **277**, 16241–16248
- Schmohl, L., Bierlmeier, J., von Kügelgen, N., Kurz, L., Reis, P., Barthels, F., Mach, P., Schutkowski, M., Freund, C., and Schwarzer, D. (2017) Identification of sortase substrates by specificity profiling. *Bioorg. Med. Chem.* **25**, 5002–5007
- Kruger, R. G., Dostal, P., and McCafferty, D. G. (2004) Development of a high-performance liquid chromatography assay and revision of kinetic parameters for the *Staphylococcus aureus* sortase transpeptidase SrtA. *Anal. Biochem.* **326**, 42–48
- Race, P. R., Bentley, M. L., Melvin, J. A., Crow, A., Hughes, R. K., Smith, W. D., Sessions, R. B., Kehoe, M. A., McCafferty, D. G., and Banfield, M. J. (2009) Crystal structure of *Streptococcus pyogenes* sortase A: Implications for sortase mechanism. *J. Biol. Chem.* **284**, 6924–6933

Selectivity determinants in class A sortases

30. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-model workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201
31. Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., and Schwede, T. (2009) Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **4**, 1–13
32. Zou, Z., Nöth, M., Jakob, F., and Schwaneberg, U. (2020) Designed *Streptococcus pyogenes* sortase A accepts branched amines as nucleophiles in sortagging. *Bioconjug. Chem.* **31**, 2476–2481
33. Chan, A. H., Yi, S. W., Terwilliger, A. L., Maresso, A. W., Jung, M. E., and Clubb, R. T. (2015) Structure of the *Bacillus anthracis* sortase A enzyme bound to its sorting signal: A flexible amino-terminal appendage modulates substrate access. *J. Biol. Chem.* **290**, 25461–25474
34. Raeeszadeh-Sarmazdeh, M., Parthasarathy, R., and Boder, E. T. (2015) Site-specific immobilization of protein layers on gold surfaces via orthogonal sortases. *Colloids Surf. B Biointerfaces* **128**, 457–463
35. Li, J., Zhang, Y., Soubias, O., Khago, D., Chao, F.-A., Li, Y., Shaw, K., and Byrd, R. A. (2020) Optimization of sortase A ligation for flexible engineering of complex protein systems. *J. Biol. Chem.* **295**, 2664–2675
36. Pritz, S., Wolf, Y., Kraetke, O., Klose, J., Bienert, M., and Beyermann, M. (2007) Synthesis of biologically active peptide nucleic acid-peptide conjugates by sortase-mediated ligation. *J. Org. Chem.* **72**, 3909–3912
37. Guimaraes, C. P., Witte, M. D., Theile, C. S., Bozkurt, G., Kundrat, L., Blom, A. E. M., and Ploegh, H. L. (2013) Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8**, 1787–1799
38. Popp, M. W., Artavanis-Tsakonas, K., and Ploegh, H. L. (2009) Substrate filtering by the active site crossover loop in UCHL3 revealed by sortagging and gain-of-function mutations. *J. Biol. Chem.* **284**, 3593–3602
39. Hess, G. T., Guimaraes, C. P., Spooner, E., Ploegh, H. L., and Belcher, A. M. (2013) Orthogonal labeling of M13 minor capsid proteins with DNA to self-assemble end-to-end multiphage structures. *ACS Synth. Biol.* **2**, 490–496
40. Wuethrich, I., Peeters, J. G. C., Blom, A. E. M., Theile, C. S., Li, Z., Spooner, E., Ploegh, H. L., and Guimaraes, C. P. (2014) Site-specific chemoenzymatic labeling of aerolysin enables the identification of new aerolysin receptors. *PLoS One* **9**, e109883
41. Valgardson, J., Cosbey, R., Houser, P., Rupp, M., Van Bronkhorst, R., Lee, M., Jagodzinski, F., and Amacher, J. F. (2019) MotifAnalyzer-PDZ: A computational program to investigate the evolution of PDZ-binding target specificity. *Protein Sci.* **28**, 2127–2143
42. Kang, C.-Y., Huang, I.-H., Chou, C.-C., Wu, T.-Y., Chang, J.-C., Hsiao, Y.-Y., Cheng, C.-H., Tsai, W.-J., Hsu, K.-C., and Wang, S. (2020) Functional analysis of *Clostridium difficile* sortase B reveals key residues for catalytic activity and substrate specificity. *J. Biol. Chem.* **295**, 3734–3745
43. Magrane, M., and Consortium, U. (2011) UniProt Knowledgebase: A hub of integrated protein data. *Database (Oxford)* **2011**, bar009
44. UniProt Consortium (2019) UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515
45. NCBI Resource Coordinators (2017) Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **45**, D12–D17
46. Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217
47. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410
48. Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004) The Jalview Java alignment editor. *Bioinformatics* **20**, 426–427
49. Johanning, K., Juliano, M. A., Juliano, L., Lazure, C., Lamango, N. S., Steiner, D. F., and Lindberg, I. (1998) Specificity of prohormone convertase 2 on proenkephalin and proenkephalin-related substrates. *J. Biol. Chem.* **273**, 22672–22680
50. Bennett, N. R., Jarvis, C. M., Alam, M. M., Zwick, D. B., Olson, J. M., Nguyen, H. V., Johnson, J. A., Cook, M. E., and Kiessling, L. L. (2019) Modular polymer antigens to optimize immunity. *Biomacromolecules* **20**, 4370–4379